Computational dissection of phenotypic and functional heterogeneity in cancer

Awardee: Dana Pe'er
Award: Pioneer Award
Awardee Institution: Columbia University
Co-authors: Jacob H. Levine, Erin F. Simonds, Sean C. Bendall and Garry P. Nolan
Co-authors' institution: Columbia University

Cells within a single tumor are known to display extensive phenotypic and functional heterogeneity. Many life-threatening features of cancer, including drug resistance, metastasis and relapse, are facets of intratumor heterogeneity. With emerging single-cell measurement technologies, the field is poised to make important strides in understanding and controlling this heterogeneity. However, these technologies require advances in analytical methods to interpret the complex data they produce.

Acute myeloid leukemia (AML) is an aggressive bone marrow malignancy in which the importance of cellular heterogeneity has been well characterized. However, previous studies have only scraped the surface of the heterogeneity in this disease. Using mass cytometry, which measures single cells in ~31 simultaneous proteomic features, we developed novel methods for analyzing phenotypic heterogeneity in cancer. Our approach provides an extensive compendium of surface-marker and signaling phenotypes in AML that extends current boundaries of knowledge.

The heart of our approach is Phenograph, a graph-based representation of single-cell samples. The graph represents the phenotypic structure of the sample and can be partitioned into subsets of densely interconnected nodes, called *communities*, which represent distinct phenotypic subpopulations. Using Phenograph, we deconstructed several AML samples into discrete phenotypes. Comparing phenotypes across patients, we found a striking degree of order. Every identifiable phenotype was discoverable in multiple (but not all) patients, implying a constraint on the space of allowable AML phenotypes.

Our data contain measurements under various environmental perturbations and we designed a method to statistically quantify evoked signaling responses, producing high-dimensional signaling phenotypes for each subpopulation, which we regard as a representation of cellular state and functional potential. We found a tight coupling between surface and signaling phenotypes in healthy cells that is disrupted in AML. We identified a primitive signaling phenotype, derived from healthy stem and progenitor cells, which was not correlated with the primitive surface marker profile typically used to define primitive cells in AML. Using single-cell frequencies to deconvolve existing bulk gene expression data, we identified genes associated with this primitive signaling phenotype. These genes produce a clinically predictive signature that is more powerful than genes associated with the primitive surface profile, validating the utility of our approach and providing a new characterization of primitive cells in AML. Phenograph can be applied to characterize heterogeneity and primitive subpopulations in additional cancers.